



Statistics Sweden

Statistiska centralbyrån

Ungdomar utan fullföljd gymnasieutbildning – en undersökning med många utmaningar

2008:1

I serien Bakgrundsfakta presenteras bakgrundsmaterial till den statistik som SCB producerar inom området arbetsmarknad och utbildning. Det kan röra sig om produktbeskrivningar, metodredovisningar samt olika sammanställningar som kan ge en överblick och underlätta användandet av statistiken.

Utgivna publikationer från 2000 i serien Bakgrundsfakta till arbetsmarknads- och utbildningsstatistiken

- 2000:1 Övergång till yrkeskodning på fyrsiffernivå (SSYK) och införande av jobbstatuskod i SCB:s lönestatistik
- 2000:2 The Information System for Occupational Injuries and the Work-related Health Problems Survey – A comparative study
- 2000:3 Konferens om utbildningsstatistik den 23 mars 2000
- 2001:1 Avvikelser i lönesummestatistiken – en jämförelse mellan LAPS och LSUM
- 2001:2 En longitudinell databas kring utbildning, inkomst och sysselsättning 1990–1998
- 2001:3 Staff training costs 1994–1999
- 2001:4 Studieresultat i högskolan i form av avklarade poäng
- 2001:5 Urvals- och estimationsförfarandet i de svenska arbetskraftsundersökningarna (AKU)
- 2001:6 Svar, bortfall och representativitet i Arbetsmiljöundersökningen 1999
- 2001:7 Individ- och företagsbaserad sysselsättningsstatistik – en jämförelse mellan AKU och KS
- 2002:1 Tidsseriebrott i utbildningsregistret 2001-01-01
- 2002:2 En longitudinell databas kring utbildning, inkomst och sysselsättning (LOUISE) 1990–1999
- 2003:1 Exempel på hur EU:s "Quality Reports" kan skrivas – avser Labour Cost Survey (LSC) 2000
- 2003:2 Förändrad redovisning av högskolans personal
- 2003:3 Individ- och företagsbaserad sysselsättningsstatistik – en fortsatt jämförelse mellan AKU och KS
- 2003:4 Sjukfrånvarande enligt SCB och sjukskrivna enligt RFV
- 2003:5 Informationssystemet om arbetsskador och undersökningen om arbetsorsakade besvär. En jämförande studie
- 2004:1 Samlad statistik från SCB avseende ohälsa
- 2004:2 Översyn av forskarutbildningsstatistiken. Bedömning av kvaliteten
- 2004:3 Sjukfrånvaro och ohälsa i Sverige – en belysning utifrån SCB:s statistik
- 2005:1 En longitudinell databas kring utbildning, inkomst och sysselsättning (LOUISE) 1990–2002
- 2005:2 Nordisk pendlingskarta. Huvudrapport
- 2005:3 Nordisk pendlingskarta. Delrapport 1–4.
- 2005:4 Flödesstatistik från AKU
- 2005:5 Flow statistics from the Swedish Labour Force Survey
- 2006:1 Sysselsättningsavgränsning i RAMS – Metodöversyn 2005

Fortsättning på omslagets tredje sida!

Ovannämnda rapporter, liksom övriga SCB-publikationer, kan beställas från:
Statistiska centralbyrån, SCB, Publikationstjänsten, 701 89 ÖREBRO,
telefon 019-17 68 00 eller fax 019-17 64 44.

Du kan också köpa SCB:s publikationer i **Statistikbutiken**:
Karlavägen 100, Stockholm

Bakgrundsfakta

**Ungdomar utan fullföljd
gymnasieutbildning
– en undersökning med många
utmaningar**

Arbetsmarknads- och utbildningsstatistik 2008:1

**Statistiska centralbyrån
2008**

Background Facts

Labour and Education Statistics 2008:1

Youths without completed upper secondary education – a survey with many challenges

Statistics Sweden
2008

Producent
Producer SCB, enheten för statistik om utbildning och arbetsmarknad
Statistics Sweden, Education and jobs
SE-701 89 ÖREBRO

Förfrågningar
Inquiries Anders Karlsson, +46 19 17 63 07
anders.karlsson@scb.se

Johan Eriksson, +46 19 17 67 38
johan.eriksson@scb.se

ua@scb.se

Det är tillåtet att kopiera och på annat sätt mångfaldiga innehållet i denna publikation.

Om du citerar, var god uppge källan på följande sätt:

Källa: SCB, Bakgrundsfakta, Arbetsmarknads- och utbildningsstatistik 2008:1, *Ungdomar utan fullföljd gymnasieutbildning – en undersökning med många utmaningar*.

It is permitted to copy and reproduce the contents in this publication.

When quoting, please state the source as follows:

Source: Statistics Sweden, Background Facts, Labour and Education Statistics 2008:1, *Youths without completed upper secondary education – a survey with many challenges*.

ISSN 1654-465X (online)

ISSN 1103-7458 (print)

ISBN 978-91-618-1430-5 (print)

URN:NBN:SE:SCB-2008-AM76BR0801_pdf (pdf)

Printed in Sweden
SCB-tryck, Örebro 2008:3

Förord

Att ungefär en fjärdedel av eleverna lämnar gymnasieskolan utan att erhålla ett slutbetyg har inneburit ett allt större intresse för studier av denna grupp. Under 2007 genomförde Statistiska centralbyrån en undersökning riktad till ungdomar som påbörjade gymnasieskolan höstterminen 2000.

Undersökningen planerades i första hand som en enkätundersökning bland dem som inte fullföljt gymnasieskolan. Så småningom kom den även att omfatta en registerstudie av samtliga ungdomar som påbörjade gymnasieskolan höstterminen 2000. Detta Bakgrundsfakta är koncentrerat till svårigheter som uppkom i samband med enkätundersökningen.

Eftersom intresset för information kring gruppen ungdomar utan fullföljd gymnasieskola är stort, har en hel del möda lagts ner för att samla in så mycket relevanta data som möjligt. Detta har varit möjligt tack vare ett starkt stöd och hjälp av metodenheten inom Avdelningen för befolkning och välfärd. Avdelningen för näringsliv och arbetsmarknad har också bidragit med sakkunskap och råd kring de register som kommit till användning i olika skeden av undersökningen.

Rapporten har utformats av Anders Karlsson och Karin Björklind vid enheten för statistik om utbildning och arbete samt Johan Eriksson vid metodenheten för individ- och hushållsstatistik.

Statistiska centralbyrån i mars 2008

Anna Wilén

Jenny Karlsson

Innehåll

Förord	3
1. Bakgrund	7
2. Genomförandet av undersökningen	9
2.1 Konsumentmöte och syftet med undersökningen	9
2.2 Formulering av frågor och frågeblanketten	9
2.2.1 Test vid SCB:s mättekniska laboratorium	9
2.3 Postenkätundersökningen	9
2.3.1 Population och avgränsningar	9
2.3.2 Urvalsdesign och urval	10
2.3.3 En utsatt grupp	10
2.3.4 Datainsamlingen	11
2.3.5 En ny strategi	11
2.4 Telefonintervjuundersökningen	11
2.4.1 Provundersökning	11
2.4.2 Urvalsdesign och urval	11
2.4.3 Datainsamlingen	12
3. Urvalsdesign	13
3.1 Slutlig urvalsdesign	13
3.2.1 Beskrivning av ett tvåfasurval	14
3.2.2 Allokering av urval fas 1	14
3.2.3 Allokering av urval fas 2	15
4. Bortfallsstudie	17
4.1 Insamlingsresultat	17
4.1.1 Rampopulationens aktualitet och kvalitet	18
4.2 Beskrivning av bortfallet	18
4.2.1 Bortfall i olika grupper	18
4.2.2 Beräkning av bortfall	19
4.2.3 Resultat av bortfallsstudie	20
5. Estimation	21
5.1 Möjliga estimatorer	21
5.2 Val av estimator	21
5.3 Val av hjälpinformation	22
5.3.1 Hjälpvariabler som var starka enligt kriterium 1	22
5.3.2 Val av hjälpinformation med hjälp av indikator	23
5.3.3 Hjälpvariabler som var starka enligt kriterium 2	24
5.3.4 Hjälpvariabler som var starka enligt kriterium 3	25
5.3.5 Val av slutlig hjälpvektor	26
5.4 Effekter av vald estimator	27
5.4.1 Effekter på skattningar	27
6. Bilagor	31
Bilaga 1. Svartsandel och bortfallsandel efter tänkbara hjälvariabler	32
Bilaga 2. Definitioner	35

1. Bakgrund

Under de senaste åren har andelen elever som fullföljt gymnasieskolan inom fem år legat mellan 74 och 77 procent. Det innebär att en fjärdedel – eller i runda tal 25 000 elever för varje årgång – lämnar gymnasieskolan utan att erhålla ett slutbetyg. Att så många inte slutför sin utbildning har till exempel lett till att regeringen i februari 2007 beslöt att tillsätta en särskild utredare med uppgift att lämna förslag till den framtida strukturen av gymnasieskolans studievägar.

Det pågår på många håll, bland annat Skolverket, Sveriges Kommuner och Landsting samt Göteborgs universitet, studier om skälen till att ungdomar avbryter gymnasieskolan och vad detta innebär för dem när det gäller etablering på arbetsmarknaden med mera. Inom SCB beslutade enheten för statistik om utbildning och arbete att inom ramen för den officiella statistiken genomföra en enkätundersökning under våren 2007, riktad till personer som påbörjat en gymnasieutbildning höstterminen 2000 men avgått utan slutbetyg. Resultat från undersökningen redovisades i publikationen Tema: Utbildning 2007 nr 4 "Ungdomar utan fullföljd gymnasieutbildning".

2. Genomförandet av undersökningen

2.1 Konsumentmöte och syftet med undersökningen

För att fånga upp användarönskemål om urval och frågor i en undersökning riktad till ungdomar utan fullföljd gymnasieutbildning inbjöds myndigheter, kommuner, landsting, större gymnasieskolor med flera till ett så kallat konsumentmöte den 12 februari 2007. 27 personer deltog vid mötet. Ytterligare några som inte kunde delta lämnade kommentarer via e-post.

Konsumentmötet och efterföljande interna diskussioner på enheten resulterade i ett beslut att syftet med undersökningen skulle vara att belysa tiden i gymnasieskolan, orsaker till studieavbrott samt etablering på arbetsmarknaden och eventuella studier efter avhoppet från gymnasieskolan. Utifrån syftet togs ett förslag på urvalsdesign och enkätfrågor fram. Innan undersökningen genomfördes gavs de inbjudna till konsumentmötet möjlighet att beställa tillägsfrågor och tilläggsurval.

2.2 Formulering av frågor och frågeblanketten

Frågorna i undersökningen formulerades av SCB i samråd med deltagarna på konsumentmötet. Frågeblanketten testades också av mättekniska laboratoriet vid SCB. Den slutliga versionen av blanketten omfattade åtta sidor (inklusive missiv) med 30 numrerade frågor, varav några var uppdelade på delfrågor.

2.2.1 Test vid SCB:s mättekniska laboratorium

Eftersom det redan vid planeringen av undersökningen fanns misstankar om att svarsbenägenheten kunde vara låg hos målgruppen bokades det mättekniska laboratoriet in på ett tidigt stadium för en test av frågeblanketten. Testet innebar inte bara en skrivbordsgranskning av blanketten, utan den prövades även på några testpersoner som tillhörde målgruppen.

Vid testet var problemet främst att etablera kontakt med personer i målgruppen. De som medverkade angav att de inte tyckte att frågorna var svåra att besvara, inte heller att frågorna upplevdes som känsliga. Språket uppfattades som enkelt och enkäten var inte för lång. Det som framförallt kom synpunkter på utifrån testet var att tänka igenom strukturen samt att ta med fler definitioner i anslutning till frågorna.

Den slutliga enkätversionen justerades utifrån mätlaboratoriets rekommendationer.

2.3 Postenkätundersökningen

2.3.1 Population och avgränsningar

Enkätundersökningens rampopulation utgjordes av nybörjare i gymnasieskolan höstterminen 2000 enligt gymnasieskolans elevregister, som inte fullföljt en gymnasieutbildning enligt gymnasieskolans register över avgångna till och med vårterminen 2006. Med fullföljd gymnasieutbildning

menades här att ett slutbetyg eller motsvarande omdöme erhöles. Totalt omfattade målgruppen 23 386 personer. Följande avgränsningar gjordes redan från början:

- personer födda före 1980 togs inte med (25 personer)
- personer som påbörjat ett individuellt program höstterminen 2000 men endast gick i gymnasieskolan denna termin togs inte med (drygt 1 000 personer). Anledningen var att många av dessa elever förmodligen bara läst in något eller några av grundskolans ämnen och inte hade för avsikt att läsa ett helt gymnasieprogram.

Under datainsamlingens gång hörde många av sig som påbörjat gymnasieprogrammet International Baccalaureate (IB) och uppgav att de hade erhållit ett slutbetyg. Av enkätsvaren angav över 60 procent från detta program samma sak. Enligt ansvariga för gymnasieskolans register över avgångna har också tidigare funnits en viss eftersläpning av rapporteringen av betygsuppgifter till SCB från skolorna med IB-programmet. Sammantaget ledde detta till beslutet att de som påbörjat IB exkluderades från undersökningen (140 personer).

Den slutliga rampopulationen kom därmed att omfatta 22 180 personer.

2.3.2 Urvalsdesign och urval

Den planerade urvalsdesignen för undersökningen var ett stratifierat obundet slumpmässigt urval. Inför urvalsdragningen stratifierades objekten i rampopulationen efter kön (två grupper), svensk/utländsk bakgrund (två grupper), påbörjat gymnasieprogram (tre grupper) och årskurs för avbrott (tre grupper).

De tillgängliga ekonomiska resurserna tillät en postenkätundersökning med en urvalsstorlek om cirka 8 000 individer, inklusive en bortfallsuppföljning med telefonintervjuer om cirka 2 000 individer. Urvalet designades så att det gav tillräcklig precision i skattningarna av de viktigaste målstorheterna.

2.3.3 En utsatt grupp

Inför utsändningen av undersökningen hämtades aktuella namn- och adressuppgifter för de utvalda personerna från registret över totalbefolkningen (RTB) vid SCB. Efter matchning mot RTB framkom att nästan en procent hade avlidit. Denna siffra motsvarar dödstaten för 70-åringar sett till hela Sveriges befolkning. Personerna i urvalet var högst 27 år gamla, 95 procent var 23-24 år.

2.3.4 Datainsamlingen

Undersökningen genomfördes med en postenkät, med möjlighet att skicka in svaren via vanlig post eller lämna dem via SCB:s webbplats.

- Enkäterna skickades ut den 19 april 2007
- Tack- och påminnelsekort sändes ut den 27 april
- Påminnelsebrev med ny enkät skickades ut den 11 maj

Vanligtvis brukar andelen svarande i en postenkätundersökning en månad efter utsändning ligga mellan 50 och 70 procent. I denna undersökning var svarsfrekvensen en vecka efter det sista påminnelseutskicket (den 17 maj) endast 20 procent.

2.3.5 En ny strategi

Det var uppenbart att en postenkät inte fungerade som insamlingsmetod för denna målgrupp. Dessutom konstaterades att bortfallet var för stort, ur både ekonomisk och tidsmässig synvinkel, för att göra en traditionell bortfallsuppföljning med telefonintervjuer.

En diskussion följde inom enhetens metodgrupp om hur vi skulle gå vidare. Ett förslag var att lägga ner undersökningen, medan ett annat var att gå vidare och slutföra undersökningen "på något sätt". Eftersom målgruppen är av så stort intresse i samhällsdebatten och eftersom informationen kring den är så bristfällig blev valet, i samråd med avdelningens metodchef, att formulera en ny strategi. Denna blev att göra en ren telefonintervjuundersökning till ett mindre urval.

2.4 Telefonintervjuundersökningen

2.4.1 Provundersökning

Innan beslut togs om att genomföra en telefonintervjuundersökning i full skala, gjordes först en provundersökning till ett urval om 60 individer. Dessa söktes av en telefonintervjuare under en veckas tid. För hälften av personerna i provundersökningen erhöles en intervju. Detta utan att något spårningsarbete gjordes för de personer som saknade eller hade ett ogiltigt telefonnummer efter matchning mot telefonoperatörernas register.

Om intervjuaren fick kontakt med de utvalda personerna genomfördes i de flesta fallen en intervju. Antalet personer som inte ville delta eller inte svarade av någon annan orsak var ytterst få. Bortfallet berodde alltså till största delen på att intervjuaren inte lyckades nå de utvalda. Antagandet om att en någorlunda lång insamlingstid tillsammans med så kallad djupspårning av telefonnummer borde ge ett acceptabelt resultat ledde därför till beslutet om att genomföra en huvudundersökning.

2.4.2 Urvalsdesign och urval

Telefonintervjuundersökningen genomfördes till ett urval om 2 500 personer, dragna ur det tidigare urvalet om 8 000 personer. Urvalet drogs proportionellt i varje ursprungligt stratum (stratifieringen efter kön exkluderas dock). Personer som tidigare svarat i postenkätundersökningen eller utgjorde känt bortfall (vägrare, långvarigt sjuka etc.) och som kom med i det

nya urvalet togs med. Innan urvalet överlämnades till telefonintervjuarna plockades de dock bort.

2.4.3 Datainsamlingen

Totalt ingick 1 785 personer i telefonintervjuundersökningen. Efter *grundspårning* hos telefonoperatörerna saknades telefonnummer för 29 procent, normalt brukar andelen vara omkring 10 procent. En *djupspårning* för att hitta telefonnummer gjordes därefter. Vid djupspårning söks telefonnummer hos skattemyndigheten, försäkringskassan, arbetsgivare, brevbärare, föräldrar med mera.

Resultatet efter djupspårningen blev att för 9 procent av alla personer i telefonintervjuundersökningen hittades inget telefonnummer alls. Normalt brukar denna andel vara omkring 6 procent, denna undersökning avviker därmed inte anmärkningsvärt från genomsnittet.

Telefonintervjuundersökningen genomfördes under maj – augusti 2007. Insamlingen avslutades under första veckan i september 2007. Det slutliga resultatet framgår av tabell 1. Intervjuer erhöles från 37 procent av urvalet. Detta är en låg andel när det gäller telefonintervjuer. Svårigheten var inte att få personer att ställa upp på en intervju, utan att hitta giltiga telefonnummer. Andelen saknade telefonnummer tillsammans med erhållna men ogiltiga telefonnummer uppgick till 27 procent för hela materialet, normalt brukar andelen vara cirka 8 procent.

Den slutliga svarsfrekvensen i undersökningen, inklusive medräknade postenkätsvar, blev 51 procent (oviktad siffra).

Tabell 1. Resultat av telefonintervjuundersökningen

Urval till telefonintervjuundersökningen	1 785 personer
Genomförda intervjuer	37%
Känt bortfall (ville ej delta, bortrest m.m.)	13%
Lovade skicka in enkät	10%
Ringt men ej antråffade	13%
Telefonnummer erhållet men ogiltigt	18%
Uppgift om telefon saknas	9%

3. Urvalsdesign

3.1 Slutlig urvalsdesign

Urvalsdesign: Stratifierat tvåfasurval

Urvalsenhet: Individ

Stratifieringsvariabler fas 1: Kön, svensk/utländsk bakgrund, påbörjat gymnasieprogram och årskurs för avbrott.

Urvalstorlek fas 1: 7 860¹

Stratifieringsvariabler fas 2: Svensk/utländsk bakgrund, påbörjat gymnasieprogram och årskurs för avbrott.

Urvalstorlek fas 2: 2 437²

Tabell 2. Stratifieringsvariabler

Kön		Svensk/utländsk bakgrund	
1	Man	1	Svensk bakgrund
2	Kvinna	2	Utländsk bakgrund
Påbörjat gymnasieprogram		Årskurs för avbrott	
1	Studieförberedande program	1	Årskurs 1
2	Yrkesförberedande program	2	Årskurs 2
3	Individuellt program	3	Årskurs 3

¹ Det ursprungliga urvalet i fas 1 var 8 000 individer. Under datainsamlingen exkluderades 140 individer som påbörjat gymnasieprogrammet International Baccalaureate (IB).

² Det ursprungliga urvalet i fas 2 var 2 500 individer. Under datainsamlingen exkluderades 63 individer som påbörjat gymnasieprogrammet International Baccalaureate (IB).

3.2.1 Beskrivning av ett tvåfasurval

En population $U = \{1, \dots, k, \dots, N\}$ bestående av N individer delas in i H strata U_h , $h=1, \dots, H$, där stratum h innehåller N_h individer och $U = \bigcup_h U_h$.

I första fasen dras ett obundet slumpmässigt urval, s_{ah} av storlek n_{ah} från stratum h . Det totala urvalet i första fasen består av n_a individer och utgörs av mängden $s_a = \bigcup_h s_{ah}$.

Första fasens urval s_a stratifieras oberoende av första fasens strata, efter G strata (grupper) i andra fasen med n_{ag} individer i grupp g , $g=1, \dots, G$.

I andra fasen dras ett obundet slumpmässigt urval, s_g av storlek n_g från varje grupp g i andra fasen, låt r_g av storlek m_g beteckna mängden av svarande individer i grupp g .

Det totala urvalet i fas 2 består av n individer och utgörs av mängden $s = \bigcup_g s_g$. Det totala antalet svarande i fas 2 består av m individer och utgörs av mängden $r = \bigcup_g r_g$.

3.2.2 Allokering av urval fas 1

Tabell 3 visar rampopulationens storlek (N_h) och urvalsstorlek (n_{ah}) fördelade efter stratum i fas 1 i undersökningen bland ungdomar utan fullföljd gymnasieutbildning.

Stratumidentiteter:

- Position 1 (1 = man, 2 = kvinna)
- Position 2 (0 = svensk bakgrund, 1 = utländsk bakgrund)
- Position 3 (1 = studieförberedande program, 2 = yrkesförberedande program, 3 = individuellt program)
- Position 4 (1 = avbrott i årskurs 1, 2 = avbrott i årskurs 2, 3 = avbrott i årskurs 3).

Urvalet allokerades för att ge tillräcklig precision i skattningarna av de viktigaste målstorheterna fördelade efter redovisningsgrupp.

Tabell 3. Stratuminformation fas 1

Stratum	N_h	n_{ah}	f_{1h}	Stratum, forts	N_h	n_{ah}	f_{1h}
1011	304	203	0,67	2011	384	222	0,58
1012	539	247	0,46	2012	482	239	0,50
1013	2 499	339	0,14	2013	1 839	326	0,18
1021	524	245	0,47	2021	547	248	0,45
1022	578	251	0,43	2022	612	255	0,42
1023	2 799	347	0,12	2023	1 889	328	0,17
1031	1 129	297	0,26	2031	839	276	0,33
1032	512	243	0,47	2032	375	220	0,59
1033	991	287	0,29	2033	535	246	0,46
1111	98	98	1,00	2111	80	80	1,00
1112	147	142	0,97	2112	124	124	1,00
1113	715	267	0,37	2113	546	248	0,45
1121	136	136	1,00	2121	125	125	1,00
1122	151	145	0,96	2122	82	82	1,00
1123	536	246	0,46	2123	361	217	0,60
1131	494	241	0,49	2131	319	207	0,65
1132	187	163	0,87	2132	127	127	1,00
1133	354	216	0,61	2133	221	177	0,80
Totalt	22 180	7 860	0,35				

där f_{1h} betecknar urvalsfraktionen i fas 1, (n_{ah}/N_h).

3.2.3 Allokering av urval fas 2

Från första fasens urval s_a om $n_a = 7\,860$ individer drogs ett stratifierat urval om $n = 2\,437$ individer. Tabell 4 visar första fasens urvalsstorlek (n_{ag}) och urvalsstorlek i fas 2 (n_g) fördelat efter stratum i fas 2.

Urvalet i fas 2 allokerades proportionellt mot första fasens urvalsstorlekar:

$$n_g = n \frac{n_{ag}}{\sum_{g=1}^G n_{ag}}$$

I och med att ett andra fasens urval drogs från första fasens urval kunde inte de precisionskrav som första fasens urval beräknades uppfylla, infrias. Urvalsstorleken i fas 2 tillät dock redovisning efter kön, svensk/utländsk bakgrund, påbörjat gymnasieprogram och årskurs för avbrott med ett konfidensintervall som inte blev "längre" än fem procentenheter (0,05). Man kunde dock inte bryta ned statistiken efter första fasens strata med bibehållen precision, beräknad enligt första fasens urval.

Stratumidentiteter:

- Position 1 (3 = både män och kvinnor, ingen stratifiering efter kön)
- Position 2 (0 = svensk bakgrund, 1 = utländsk bakgrund)
- Position 3 (1 = studieförberedande program, 2 = yrkesförberedande program, 3 = individuellt program)
- Position 4 (1 = avbrott i årskurs 1, 2 = avbrott i årskurs 2, 3 = avbrott i årskurs 3)

Tabell 4. Stratuminformation fas 2

Stratum	n_{ag}	n_g	f_{2g}	Stratum, forts	n_{ag}	n_g	f_{2g}
3011	425	132	0,31	3111	178	56	0,31
3012	486	151	0,31	3112	266	82	0,31
3013	665	206	0,31	3113	515	160	0,31
3021	493	153	0,31	3121	261	81	0,31
3022	506	157	0,31	3122	227	70	0,31
3023	675	209	0,31	3123	463	143	0,31
3031	573	178	0,31	3131	448	139	0,31
3032	463	143	0,31	3132	290	90	0,31
3033	533	165	0,31	3133	393	122	0,31
				Totalt	7 860	2 437	0,31

där f_{2g} betecknar urvalsfraktionen i fas 2, (n_g/n_{ag}) .

4. Bortfallsstudie

4.1 Insamlingsresultat

Av de utvalda 2 437 personerna var det 1 240 individer som svarade via telefonintervju eller inskickad postenkät. Detta motsvarade 51 procent av urvalet.

Andel svarande har i tabell 5 beräknats som (m_g/n_g) . Dock fanns det individer i urvalet som inte ansågs tillhöra målpopulationen (se avsnitt 4.2.1).

Vid beräkning av svarsfrekvenser ska enligt SCB:s standard tas hänsyn till designvikter och övertäckning. Tabell 5 visar hur stor andel av urvalet (inklusive övertäckning) som svarat och ska därför inte ses som ett mått på svarsfrekvensen. I bilaga 1 presenteras svarsfrekvenser beräknade enligt SCB:s standard.

Tabell 5. Antal och andel svarande individer per stratum, fas 2

Stratum2	n_{ag}	n_g	m_g	Andel svar i %
3011	425	132	69	52
3012	486	151	92	61
3013	665	206	124	60
3021	493	153	75	49
3022	506	157	88	56
3023	675	209	122	58
3031	573	178	77	43
3032	463	143	84	59
3033	533	165	89	54
3111	178	56	17	30
3112	266	82	41	50
3113	515	160	66	41
3121	261	81	36	44
3122	227	70	35	50
3123	463	143	73	51
3131	448	139	51	37
3132	290	90	46	51
3133	393	122	55	45
Totalt	7 860	2 437	1 240	51

4.1.1 Rampopulationens aktualitet och kvalitet

Målpopulationen i undersökningen utgjordes av nybörjare i gymnasieskolan höstterminen 2000 enligt gymnasieskolans elevregister, som inte fullföljt en gymnasieutbildning enligt gymnasieskolans register över avgångna.

En rampopulation bestående av 22 180 individer skapades från gymnasieskolans elevregister.

Vid matchning av rampopulationens 22 180 individer mot en aktuell version av RTB visade det sig att 675 individer inte återfanns i RTB. Den vanligaste anledningen till att en individ i rampopulationen inte matchade mot RTB var att individen flyttat utomlands mellan tiden för påbörjade gymnasiestudier och 2007-09-30. Det var 295 individer som hade flyttat utomlands. En annan anledning var att individen avlidit (161 individer) eller att personnumret i gymnasieskolans register över avgångna var felaktigt (209 individer).

De 161 individer som avlidit ansågs inte tillhöra målpopulationen. Individer som flyttat utomlands och individer med felaktiga personnummer ansågs tillhöra målpopulationen. Dessa har vid beräkning av svarsfrekvenser behandlats som bortfall (se avsnitt 4.2.2).

De resultat som publicerats från undersökningen omfattar endast de individer som matchade mot RTB. Anledningen till att individer som inte matchade mot RTB uteslutits ur resultaten är att svar saknades från denna grupp och att även registerinformation i stor utsträckning saknades.

Av urvalets 2 437 individer var det 109 individer som inte matchade mot RTB (version 2007-09-30). Från dessa individer har inget svar erhållits. Av de 109 individerna var det 18 som avlidit och inte ansågs tillhöra målpopulationen.

4.2 Beskrivning av bortfallet

4.2.1 Bortfall i olika grupper

På grund av den låga andelen svarande i undersökningen var det av stort intresse att jämföra svarande individer med bortfallet med avseende på bakgrundsvariabler. Avvek bortfallet från de svarande med avseende på information som finns tillgänglig i olika register?

Eftersom populationen bestod av unga personer födda efter 1979 som inte fullföljt gymnasieskolan var en del av de variabler som normalt används, exempelvis ålder, civilstånd och utbildningsnivå, troligen inte intressanta. Det fanns därför anledning att inventera vilka variabler som kunde användas.

Nedan listas de variabler som användes för att studera bortfallet. För definitioner och klassindelningar av variablerna hänvisas till bilaga 2.

Tabell 6. Variabler som användes för att studera bortfallet

Variabel	Källa
Meritvärde årskurs 9	Grundskolans årskurs 9-register, vårterminerna 1996-2000
Region	RTB, version 2007-09-30
Förvärvsinkomst	Inkomst- och taxeringsregistret, 2005
Studerande	Registret över personer i utbildning, 2005
Kön	Gymnasieskolans elevregister, 2000
Sysselsättningsstatus	Sysselsättningsregistret, 2005
Svensk/utländsk bakgrund	Gymnasieskolans elevregister, 2000
Socialbidrag	LISA-databasen, 2005
Påbörjat gymnasieprogram	Gymnasieskolans elevregister, 2000-2006
Årskurs för avbrott	Gymnasieskolans elevregister, 2000
Arbetsmarknadspolitisk åtgärd	LISA-databasen, 2005
Arbetslöshetsersättning	LISA-databasen, 2005
Bostadsbidrag	LISA-databasen, 2005
Föräldrarnas utbildningsnivå	Utbildningsregistret, 2000-2005
Status	Se bilaga 3
Sjukersättning	LISA-databasen, 2005
Förtidspension	LISA-databasen, 2005
Värnpliktsersättning	LISA-databasen, 2005

4.2.2 Beräkning av bortfall

I bilaga 1 visas antal individer i urvalet efter bakgrundsvariablerna listade i tabell 6. Dessutom visas vägd svarsandel och vägd bortfallsandel. Vägd svarsandel och vägd bortfallsandel beräknades enligt följande:

$$\text{Vägd svarsandel: } \frac{100 \sum_r d_k}{\sum_r d_k + \sum_b d_k} \quad (4.1)$$

$$\text{Vägd bortfallsandel: } \frac{100 \sum_b d_k}{\sum_r d_k + \sum_b d_k} \quad (4.2)$$

där

r betecknar mängden av individer som svarat

b betecknar mängden av individer tillhörande målpopulationen som inte svarat

och

$$d_k = \frac{N_h}{n_{ah}} \frac{n_{ag}}{n_g}, \text{ betecknar designvikten för individ } k. \quad (4.3)$$

Individer med okänd målpopulationsstatus, exempelvis "ej avhörda" har antagits tillhöra målpopulationen.

Individer som inte matchade mot RTB, men som ansågs tillhöra målpopulationen, har definierats som en egen kategori vid beräkning av bortfallet. Detta med anledning av att registerinformation i stor utsträckning saknades för denna grupp.

4.2.3 Resultat av bortfallsstudie

Tabellerna i bilaga 1 visar att:

- Individer boende i storstäder och förortskommuner svarade i lägre grad än individer boende i övriga delar av Sverige.
- Individer som erhöll socialbidrag under 2005 svarade i lägre utsträckning än individer som inte erhöll socialbidrag under 2005.
- Individer som påbörjade sina gymnasiestudier på det individuella programmet svarade i lägre utsträckning än individer som påbörjade sina studier på yrkesförberedande eller studieförberedande program.
- Individer med högt betygsmeritvärde från grundskolans årskurs 9 svarade i högre grad än individer med lågt meritvärde.
- Individer som avbröt sina gymnasiestudier i årskurs 1 svarade i lägre grad än individer som avbröt sina gymnasiestudier i årskurs 2 eller årskurs 3.
- Individer med lågutbildade föräldrar (ingen av föräldrarna hade högre utbildning än grundskola) svarade i lägre grad än individer där minst en av föräldrarna hade gymnasial utbildning eller högre.
- Individer med status "Utanför arbetsmarknaden" svarade i lägre grad än individer med annan status (studerande, värnpliktig eller i någon grad etablerad på arbetsmarknaden).
- Individer som erhöll sjukersättning under 2005 svarade i lägre grad än individer som inte erhöll sjukersättning under 2005.
- Individer med hög inkomst under 2005 svarade i högre grad än individer med ingen eller låg inkomst under 2005.
- Sysselsatta och ej sysselsatta med kontrolluppgift under 2005 svarade i högre grad än ej sysselsatta utan kontrolluppgift.
- Individer med svensk bakgrund svarade i högre grad än individer med utländsk bakgrund.
- Individer som erhöll värnpliktsersättning under 2005 svarade i högre utsträckning än individer som inte erhöll värnpliktsersättning under 2005.

Även för andra variabler, exempelvis kön, studerande och förtidspension, fanns det skillnader i svarsfrekvens mellan olika grupper. Skillnaderna var dock små.

5. Estimation

I och med att bortfallet i undersökningen var högt och att tabellen i bilaga 1 indikerar att svarsbeteendet skilde sig åt mellan olika grupper fanns det anledning av att fundera på vilken estimator som skulle användas.

5.1 Möjliga estimatorer

Estimator 1 - HT-estimation av en total

Under antagande att svarsmängden, r_g är ett obundet slumpmässigt urval från andra fasens urval, $s_{g'}$ ges HT-estimatorn av en total $t_y = \sum_U y_k$, av

$$\hat{t}_y = \sum_h \frac{N_h}{n_{ah}} \sum_g \frac{n_{ag}}{m_g} \sum_{rg} y_k = \sum_h w_{1k} \sum_g w_{2k} \sum_{rg} y_k = \sum_r w_k^* y_k \quad (5.1)$$

$$\text{där } w_k^* = w_{1k} w_{2k} = \frac{N_h n_{ag}}{n_{ah} m_g} \quad (5.2)$$

Denna estimator utnyttjar ingen hjälpinformation utöver den som används i urvalsdesignen.

Estimator 2 - GREG-estimation av en total

$$\hat{t}_{ry} = \sum_r w_k^* g_k^* y_k = \sum_r \hat{w}_k^* y_k \quad (5.3)$$

där

w_k^* ges av (5.2)

och g_k^* betecknar vikt som bestäms av den använda modellen för bortfalls-korrigerig. Estimatorn utnyttjar hjälpinformation.

För formeluttryck för g_k^* hänvisas till "TOPECA – User's guide" av Claes Andersson (sidan 11).

5.2 Val av estimator

Om bortfallet avviker från de svarande med avseende på målvariablerna kan statistiken beräknad från de svarande vara behäftad med systematiska fel. Detta fel kallas för *bortfallsfel*. Bortfallsfelet kan reduceras genom att utnyttja hjälpinformation (tillgänglig via register) vid estimationen. Därför valdes i denna undersökning estimator 2 (GREG-estimatorn). Den valda estimatorn utnyttjar hjälpinformation i båda faserna.

Förutom skattningar av totaler som ges av (5.3) ville vi också kunna skatta variansen för vald estimator. För formeluttryck för variansskattningar hänvisas till "TOPECA – User's guide" av Claes Andersson (sidorna 8-9).

5.3 Val av hjälpinformation

Efter att ha valt estimator 2 valdes de variabler som skulle användas i estimationen. Vi önskade bilda en hjälpvektor \mathbf{x}_k , av mängden tillgängliga registervariabler, som så väl som möjligt reducerade bortfallsfelet. Vid val av hjälpvariabler är det tre kriterier som ska beaktas (se "Estimation in the presence of non-response (2001)" av Sixten Lundström och Carl-Erik Särndal):

- 1) Det första kriteriet är att variabeln samvarierar väl med svarsbenägenheten (-sannolikheten). Det är det viktigaste kriteriet eftersom det leder till en minskning av bortfallskevheter för alla skattningar.
- 2) Det andra kriteriet är att variabeln samvarierar väl med (viktiga) målvariabler. Om så är fallet minskar bortfallsbiasen för de skattningar som byggs upp av dessa målvariabler. Även variansen minskar för dessa skattningar.
- 3) Det tredje kriteriet är att variabeln avgränsar (viktiga) redovisningsgrupper. Det leder framförallt till minskad varians i skattningar för dessa redovisningsgrupper.

5.3.1 Hjälpvariabler som var starka enligt kriterium 1

Tabellerna i bilaga 1 indikerar vilka variabler som var aktuella enligt kriterium 1. Tabellerna är dock svåranalyserade då de flesta variabler uppvisade skillnader i svarsandelar mellan olika kategorier. Skillnaderna var dock i många fall inte så stora att det var givet vilka variabler som borde användas.

Meritvärde, region och sysselsättningsstatus var bland de variabler som uppvisade störst skillnader men i stort sett alla variabler var presumtiva hjälpvariabler.

För att få ytterligare information om vilka variabler som var starka hjälpvariabler enligt kriterium 1 användes en indikator vilken beskrivs i nästa avsnitt.

5.3.2 Val av hjälpinformation med hjälp av indikator

Indikatorn som utnyttjades i det följande är ett hjälpmedel för att välja ut variabler som uppfyller kriterium 1. Indikatorn beskriven i "Estimation in surveys with non-response (2007)" av Sixten Lundström och Carl-Erik Särndal, har följande utseende:

$$\hat{Q} = \frac{1}{\sum_r d_k} \sum_r d_k (m_k - \bar{m}_{r;d})^2 \quad (5.4)$$

där

$$m_k = \left(\sum_s d_k \mathbf{x}_k \right)' \left(\sum_r d_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k, \quad (5.5)$$

$$\bar{m}_{r;d} = \frac{\sum_r d_k m_k}{\sum_r d_k} \quad (5.6)$$

och d_k definieras i (4.3).

Ju större värde på \hat{Q} desto mer kan vi förvänta oss att bortfallsbiasen reduceras. Vid beräkningarna uteslöts de 109 individer som inte återfanns i RTB. För dessa individer saknades till stor del information avseende de aktuella variablerna. Dessutom uteslöts variablerna förtidspension, sjukersättning och värnpliktersättning. Detta med anledning av att det är få individer i urvalet som erhållit dessa former av ersättningar.

Lundström och Särndal beskriver ett datorprogram (*\hat{Q} -Forward Selection*) som utför en stegvis selektion av hjälpvariabler från poolen av registervariabler. I första steget väljs den hjälpvariabel som ger största värdet på indikatorn. Därefter bildas lika många vektorer som antalet resterande variabler. Varje sådan vektor består av den utvalda hjälpvariabeln i första steget "plus" en resterande variabel. Indikatorvärdet beräknas för alla dessa vektorer och den vektor som ger största indikatorvärdet väljs. Proceduren fortsätter därefter med de registervariabler som då återstår.

De största förändringarna av indikatorvärdena sker i början av proceduren och blir därefter mindre och mindre. Indikatorn ger inget tydligt svar på när hjälpvektorn är tillräckligt bra. Man måste också tänka på kriterierna 2 och 3 liksom hur hjälpvektorn påverkar variansen. När det finns enstaka avvikande vikter kan variansen bli stor och variansestimern till och med urarta. Även punktskattningen kan påverkas när en avvikande vikt avser ett objekt i en liten redovisningsgrupp. Det krävs alltså en bedömning där alla dessa faktorer beaktas.

Tabell 7. \hat{Q} -Forward Selection utförd på 15 presumtiva hjälpvariabler

Step	Adderad hjälpvariabel	Värde på $1000 \times \hat{Q}$
0	-----	0
1	Region	58
2	Meritvärde årskurs 9	128
3	Sysselsättningsstatus	180
4	Föräldrarnas utbildningsnivå	194
5	Studerande	204
6	Årskurs för avbrott	213
7	Socialbidrag	219
8	Status	226
9	Påbörjat gymnasieprogram	231
10	Förvärvsinkomst	235
11	Kön	239
12	Svensk/utländsk bakgrund	242
13	Arbetsmarknadspolitisk åtgärd	244
14	Bostadsbidrag	245
15	Arbetslöshetsersättning	245

Resultatet indikerade att region, meritvärde och sysselsättningsstatus var de starkaste hjälpvariablerna. En tänkbar hjälpvektor skulle kunna innehålla variabel 1-12, men det slutliga valet av hjälpvektor borde också ta hänsyn till kriterierna 2 och 3.

5.3.3 Hjälpvariabler som var starka enligt kriterium 2

För att få en uppfattning om vilka variabler som var starka enligt kriterium 2 studerades samvariationen med en viktig målvariabel. Den målvariabel som användes var om individen varit arbetslös, sedan avbrottet skedde. Frågan i enkäten löd "Har du efter gymnasieskolans slut varit arbetslös under någon sammanhängande period längre än 3 månader?"

Andelen som varit arbetslösa skattas med hjälp av vikten w_k^* som definieras i (5.2).

$$\hat{P}_{yD} = 100 \frac{\sum_r w_k^* y_k I_k}{\sum_r w_k^* I_k} \quad (5.7)$$

där $y_k = 1$ om individ k varit arbetslös minst tre månader, 0 annars

och $I_k = 1$ om individ k tillhör grupp D , 0 annars

Tabell 8 indikerar stora skillnader i andelen arbetslösa mellan olika kategorier. Framför allt förvärvsinkomst, sysselsättningsstatus, socialbidrag, arbetsmarknadspolitisk åtgärd och status borde vara starka hjälpvariabler enligt kriterium 2.

Tabell 8. Skattning av andel arbetslösa minst tre månader, procent

Variabel / Kategorier	%	Variabel / Kategorier	%
Totalt	51		
Arbetsmarknadspolitisk åtgärd		Region	
Ja	81	Storstäder + förortskommuner	43
Nej	45	Större och medelstora städer	56
		Övriga kommuner	51
Arbetslöshetsersättning		Socialbidrag	
Ja	65	Ja	71
Nej	47	Nej	44
Bostadsbidrag		Status	
Ja	59	Studerande eller värnpliktig	52
Nej	49	Etablerad	17
Föräldrarnas utbildningsnivå		Osäker ställning	43
Grundskola eller lägre	49	Svag ställning	69
Gymnasial utbildning	52	Utanför arbetsmarknaden	51
Eftergymnasial utbildning	49	Studerande	
Förvärvsinkomst		Studerat 2005	54
- 60 000	65	Ej studerat 2005	49
60 001 – 97 000	64	Svensk/utländsk bakgrund	
97 001 – 134 000	44	Svensk bakgrund	51
134 001 -	26	Utländsk bakgrund	50
Kön		Sysselsättningsstatus	
Kvinna	53	Sysselsatt	36
Man	50	Ej sysselsatt, med kontrolluppgift	69
Meritvärde årskurs 9		Ej sysselsatt, ej kontrolluppgift	63
-100 eller saknas	59	Årskurs för avbrott	
105-160	52	Årskurs 1	58
165-185	50	Årskurs 2	56
190-320	43	Årskurs 3	47
Påbörjat gymnasieprogram			
Studieförberedande program	50		
Yrkesförberedande program	46		
Individuella programmet	60		

5.3.4 Hjälpvariabler som var starka enligt kriterium 3

Variablerna kön, årskurs för avbrott, påbörjat gymnasieprogram och svensk/utländsk bakgrund avgränsade viktiga redovisningsgrupper och borde därför vara med i hjälpvektorn.

5.3.5 Val av slutlig hjälpvektor

I den slutliga hjälpvektorn användes nio hjälpvariabler.

Studering togs inte med i vektorn då den var svag avseende kriterium 2. De studerande fångades också upp av variabeln status.

Förvärvsinkomst var stark avseende kriterium 2 men inkomst fanns med i variabeln status, därför valdes den variabeln istället. Status var den variabel som bedömdes som starkast enligt kriterium 2 (stora skillnader i andelen som varit arbetslös mellan grupperna).

Variablerna socialbidrag och arbetsmarknadspolitisk åtgärd togs inte heller med i hjälpvektorn trots att de var starka avseende kriterium 2. Dessa fångades till stor del upp av variabeln status.

Slutlig hjälpvektor:

- Region (kriterium 1)
- Meritvärde årskurs 9 (kriterium 1)
- Sysselsättningsstatus (kriterium 1 och 2)
- Föräldrarnas utbildningsnivå (kriterium 1 och 2)
- Årskurs för avbrott (kriterium 3)
- Status (kriterium 2)
- Påbörjat gymnasieprogram (kriterium 3)
- Kön (kriterium 3)
- Svensk/utländsk bakgrund (kriterium 3)

Denna vektor uppvisade inga avvikande stora vikter (eller negativa) och torde därför inte störa variansskattningarna.

5.4 Effekter av vald estimator

I detta avsnitt studeras effekten av vald estimator jämfört med en HT-estimator. För några viktiga målvariabler och redovisningsgrupper studerades i vilken riktning skattningarna förändrades. Från dessa studier bedömdes om bortfallsfelet minskades med vald estimator jämfört med en HT-estimator.

De målvariabler som studerades var:

- 1) Andel individer med en förvärvsinkomst över 100 000 kronor under 2005
- 2) Andel individer som varit arbetslösa minst tre månader efter att avbrottet skedde
- 3) Andel individer som arbetade under mätveckan

I tabell 9-11 presenteras andelsskattningar med skattning av estimatorns standardavvikelse inom parentes.

5.4.1 Effekter på skattningar

Först skattades andelen individer som under 2005 hade en förvärvsinkomst över 100 000 kronor. Denna uppgift hämtades från register och var därmed tillgänglig för hela urvalet. Skattningar från hela urvalet jämfördes med HT-skattning (5.1) och GREG-skattning (5.3) baserad på de svarande.

Tabell 9. Skattningar av målvariabel 1 med olika estimatorer

Målvariabel: Andel med förvärvsinkomst 2005 över 100 000 kr samt estimatorns standardavvikelse, procent			
	HT-skattning från urval	HT-skattning från svarande	GREG-skattning från svarande
Totalt	66,6 (1,07)	68,2 (1,45)	66,2 (0,78)
Kön			
Kvinnor	62,6 (1,57)	64,1 (2,08)	61,5 (1,94)
Män	69,6 (1,45)	71,5 (2,00)	69,6 (1,47)
Påbörjat gymnasieprogram			
Studieförberedande	67,6 (1,97)	70,0 (2,54)	67,9 (2,28)
Yrkesinriktat	73,9 (1,75)	75,1 (2,31)	73,3 (2,04)
Individuella programmet	55,5 (1,73)	56,5 (2,62)	54,4 (2,43)
Årskurs för avbrott			
Årskurs 1	52,0 (1,93)	54,6 (2,91)	52,1 (2,77)
Årskurs 2	59,4 (1,81)	61,6 (2,44)	59,1 (2,61)
Årskurs 3	74,0 (1,52)	75,2 (2,00)	73,3 (1,52)
Svensk/utländsk bakgrund			
Svensk bakgrund	69,1 (1,29)	69,4 (1,73)	67,5 (1,09)
Utländsk bakgrund	57,4 (1,55)	63,9 (2,35)	61,4 (2,42)

Det var uppenbart att en HT-skattning från svarande överskattade andelen individer med en förvärvsinkomst över 100 000 kr. Detta berodde på att de med höga inkomster svarade i högre grad än de med låga inkomster.

När GREG-estimatoren användes överensstämde skattningarna väl med de skattningar som erhöles från hela svars mängden. Undantaget var för gruppen med utländsk bakgrund där en överskattning fortfarande förelåg. Den var dock betydligt mindre jämfört med en HT-estimator. GREG-estimatoren hade en dämpande effekt på överskattningen.

Därefter skattades andelen individer som varit arbetslösa (minst tre månader), sedan avbrottet skedde. Denna uppgift hämtades från enkäten, vilket medförde att information bara finns för de svarande.

Teorin var att de som varit arbetslösa var underrepresenterade bland de svarande, d.v.s. andelen arbetslösa underskattades med HT-estimatoren.

Tabell 10. Skattningar av målvariabel 2 med olika estimatorer

Målvariabel: Andel arbetslösa minst tre månader samt estimatorns standardavvikelse, procent		
	HT-skattning från svarande	GREG-skattning från svarande
Totalt	50,9 (1,61)	51,7 (1,50)
Kön		
Kvinnor	52,6 (2,23)	54,0 (2,22)
Män	49,5 (2,29)	50,0 (2,13)
Påbörjat gymnasieprogram		
Studieförberedande	49,8 (2,84)	50,6 (2,83)
Yrkesinriktat	45,7 (2,76)	46,5 (2,65)
Individuella programmet	59,5 (2,60)	60,2 (2,58)
Årskurs för avbrott		
Årskurs 1	57,7 (2,89)	58,2 (2,94)
Årskurs 2	56,4 (2,48)	56,9 (2,54)
Årskurs 3	46,7 (2,34)	47,9 (2,19)
Svensk/utländsk bakgrund		
Svensk bakgrund	51,1 (1,93)	51,7 (1,80)
Utländsk bakgrund	50,2 (2,55)	51,7 (2,59)

GREG-skattningen var överlag något högre än HT-skattningen. Om det var så att en negativ bias förelåg så minskade GREG-estimatorn denna bias. Skillnaden mellan de två estimatorerna var dock relativt liten. Det är inte möjligt att uttala sig om vilken bias som kvarstår.

Avslutningsvis skattades andelen individer som arbetade under mätveckan. Denna uppgift hämtades från enkäten. Det var troligt att individer som arbetade under mätveckan var överrepresenterade bland de svarande. Bilaga 1 visade att de som var sysselsatta under 2005 svarade i högre grad än de som inte var sysselsatta under 2005. Det var därför rimligt att anta att detta gällde även för mätveckan.

Tabell 11. Skattningar av målvariabel 3 med olika estimatorer

Målvariabel: Andel som arbetade under mätveckan samt estimatorns standardavvikelse, procent		
	HT-skattning från svarande	GREG-skattning från svarande
Totalt	59,5 (1,55)	59,1 (1,42)
Kön		
Kvinnor	48,8 (2,23)	48,4 (2,24)
Män	67,8 (2,11)	66,9 (1,96)
Påbörjat gymnasieprogram		
Studieförberedande	55,1 (2,79)	54,7 (2,75)
Yrkesinriktat	65,8 (2,53)	66,0 (2,44)
Individuella programmet	56,3 (2,63)	55,0 (2,63)
Årskurs för avbrott		
Årskurs 1	53,7 (2,92)	52,3 (2,96)
Årskurs 2	50,9 (2,47)	50,5 (2,53)
Årskurs 3	64,1 (2,21)	64,0 (2,08)
Svensk/utländsk bakgrund		
Svensk bakgrund	59,4 (1,85)	59,0 (1,71)
Utländsk bakgrund	60,0 (2,47)	59,3 (2,52)

Om en positiv bias förelåg så minskades den med den valda GREG-estimatorn. Skillnaden mellan de två estimatorerna var dock relativt liten.

Sammanfattningsvis:

- Överskattningen av andelen individer med inkomst över 100 000 kr (avser 2005) minskade med vald estimator.
- Underskattningen av andelen individer som varit arbetslösa sedan avbrottet skedde minskade troligen med vald estimator.
- Överskattningen av andelen individer som arbetade under mätveckan minskade troligen med vald estimator.
- Det är inte möjligt att uttala sig om kvarstående bias för de variabler som hämtades från enkäten.
- Den skattade standardavvikelsen minskade överlag med vald estimator jämfört med HT-estimatorn.

6. Bilagor

Bilaga 1. Svarsandel och bortfallsandel efter tänkbara hjälpvariabler

Svarsandel och bortfallsandel efter tänkbara hjälpvariabler

Variabel / Kategorier	Antal i urvalet	Vägd svarsandel	Vägd bortfallsandel
Hela urvalet	2 419	53,6	46,4
därav ej i RTB	91	0	100
Arbetslöshetsersättning			
Ja	406	58,6	41,4
Nej	1 922	54,0	46,0
Ej i RTB	91	0	100
Arbetsmarknadspolitisk åtgärd			
Ja	374	55,7	44,3
Nej	1 954	54,8	45,2
Ej i RTB	91	0	100
Bostadsbidrag			
Ja	582	50,4	49,6
Nej	1 746	56,1	43,9
Ej i RTB	91	0	100
Förtidspension			
Ja	78	51,1	48,9
Nej	2 250	55,0	45,0
Ej i RTB	91	0	100
Förvärvsinkomst			
- 60 000	747	53,7	46,3
60 001 – 97 000	663	51,4	48,6
97 001 – 134 000	510	56,8	43,2
134 001 -	408	59,3	40,7
Ej i RTB	91	0	100
Föräldrarnas utbildningsnivå			
Grundskola eller lägre	533	50,1	49,9
Gymnasial	1 206	55,8	44,2
Eftergymnasial	589	56,3	43,7
Ej i RTB	91	0	100
Meritvärde			
-100 eller saknas	726	45,3	54,7
105-160	613	56,2	43,8
165-185	499	54,0	46,0
190-320	490	64,1	35,9
Ej i RTB	91	0	100
Kön			
Kvinna	1 029	57,1	42,9
Man	1 229	53,4	46,6
Ej i RTB	91	0	100

Svarsandel och bortfallsandel efter tänkbara hjälpvariabler, forts

Variabel / Kategorier	Antal i urvalet	Vägd svarsandel	Vägd bortfallsandel
Hela urvalet	2 419	53,6	46,4
därav ej i RTB	91	0	100
Region			
Storstäder + Förortskommuner	832	47,6	52,4
Större och medelstora städer	1 009	58,4	41,6
Övriga kommuner	487	58,7	41,3
Ej i RTB	91	0	100
Påbörjat gymnasieprogram			
Studieförberedande program	733	58,0	42,0
Yrkesförberedande program	781	57,0	43,0
Individuella programmet	814	48,3	51,7
Ej i RTB	91	0	100
Sjukersättning			
Ja	138	47,1	52,9
Nej	2 190	55,5	44,5
Ej i RTB	91	0	100
Status			
Studerande eller värnpliktig	796	57,5	42,5
Etablerad	359	56,0	44,0
Osäker ställning	219	59,9	40,1
Svag ställning	600	53,6	46,4
Utanför arbetsmarknaden	354	44,8	55,2
Ej i RTB	91	0	100
Studerande			
Studerat 2005	853	57,6	42,4
Ej studerat 2005	1 475	53,5	46,5
Ej i RTB	91	0	100
Socialbidrag			
Ja	792	47,3	52,7
Nej	1 536	57,9	42,1
Ej i RTB	91	0	100
Svensk/utländsk bakgrund			
Svensk bakgrund	1 441	57,2	42,8
Utländsk bakgrund	887	46,4	53,6
Ej i RTB	91	0	100

Svarsandel och bortfallsandel efter tänkbara hjälpvariabler, forts

Variabel / Kategorier	Antal i urvalet	Vägd svarsandel	Vägd bortfallsandel
Hela urvalet	2 419	53,6	46,4
därav ej i RTB	91	0	100
Sysselsättningsstatus			
Sysselsatt	1 010	57,6	42,4
Ej sysselsatt, med kontrolluppgift	643	56,7	43,3
Ej sysselsatt, ej kontrolluppgift	675	47,7	52,3
Ej i RTB	91	0	100
Värnpliktsersättning			
Ja	53	63,7	36,3
Nej	2 275	54,7	45,3
Ej i RTB	91	0	100
Årskurs för avbrott			
Årskurs 1	677	46,9	53,1
Årskurs 2	670	58,0	42,0
Årskurs 3	981	56,9	43,1
Ej i RTB	91	0	100

Bilaga 2. Definitioner

Variabel	Definition
Arbetslöshetsersättning	Inkomst föranledd av arbetslöshet. 0 = Ingen inkomst av detta slag (eller uppgift saknas) 1 = Inkomst av detta slag (minst 100 kr)
Arbetsmarknadspolitisk åtgärd	Inkomst föranledd av arbetsmarknadspolitisk åtgärd. 0 = Ingen inkomst av detta slag (eller uppgift saknas) 1 = Inkomst av detta slag (minst 100 kr)
Bostadsbidrag	Inkomst av bostadsbidrag finns i hushållet. 0 = Ingen inkomst av detta slag (eller uppgift saknas) 1 = Inkomst av detta slag (minst 100 kr)
Förtidspension	Inkomst föranledd av förtidspension/sjukbidrag. 0 = Ingen inkomst av detta slag (eller uppgift saknas) 1 = Inkomst av detta slag (minst 100 kr)
Förvärvsinkomst	Inkomst av förvärvskälla. Kontant bruttolön och inkomst av aktiv näringsverksamhet. 1 = 0 – 60 000 kr (eller uppgift saknas) 2 = 60 001 – 97 000 kr 3 = 97 001 – 134 000 kr 4 = 134 001 kr –
Föräldrarnas utbildningsnivå	1 = Ingen av föräldrarna har högre utbildning än grundskola (eller uppgift saknas) 2 = Minst en av föräldrarna har gymnasial utbildning 3 = Minst en av föräldrarna har eftergymnasial utbildning
Kön	1 = Man 2 = Kvinna
Meritvärde årskurs 9	Meritvärdet utgörs av summan av betygsvärdena för de 16 bästa betygen i elevens slutbetyg (G=10, VG=15 och MVG=20). 1 = -100 (eller uppgift saknas) 2 = 105-160 3 = 165-185 4 = 190-320

Variabel	Definition
Region	<p>Sveriges Kommuner och Landsting delar in kommunerna i nio grupper efter strukturella egenskaper som till exempel befolkningsstorlek, pendlingsmönster och näringslivsstruktur.</p> <p>Storstad (Stockholm, Göteborg, Malmö) Kommun med en folkmängd som överstiger 200 000 invånare.</p> <p>Förortskommun (36 kommuner) Kommun där mer än 50 procent av nattbefolkningen pendlar till arbetet i någon annan kommun. Det vanligaste utpendlingsmålet skall vara en storstad.</p> <p>Större stad (26 kommuner) Kommun med 50 000–200 000 invånare samt med mindre än 40 procent av nattbefolkningen sysselsatta inom industrisektorn.</p> <p>Medelstor stad (40 kommuner) Kommun med 20 000–50 000 invånare, med tätortsgrad över 70 procent samt med mindre än 40 procent av nattbefolkningen inom industrisektorn.</p> <p>Industrikommun (53 kommuner) Kommun med mer än 40 procent av nattbefolkningen sysselsatta inom industrisektorn och som inte är glesbygdskommun.</p> <p>Landsbygdskommun (30 kommuner) Kommun med mer än 6,4 procent av nattbefolkningen sysselsatta inom jord- och skogssektorn och som inte är glesbygdskommun.</p> <p>Glesbygdskommun (29 kommuner) Kommun med mindre än 5 invånare per kvadratkilometer och mindre än 20 000 invånare.</p> <p>Övrig större kommun (31 kommuner) Övrig kommun med 15 000–50 000 invånare.</p> <p>Övrig mindre kommun (42 kommuner) Övrig kommun med mindre än 15 000 invånare.</p> <p>I denna rapport har en sammanslagning av dessa grupper gjorts enligt</p> <p>1 = Storstäder och förortskommuner 2 = Större städer och medelstora städer 3 = Övriga kommuner</p>
Påbörjat gymnasieprogram	<p>Påbörjat gymnasieprogram bland nybörjare i gymnasieskolan höstterminen 2000.</p> <p>1 = Studieförberedande program 2 = Yrkesförberedande program 3 = Individuellt program</p>
Sjukersättning	<p>Inkomst föranledd av sjukdom/arbetskada/rehabilitering.</p> <p>0 = Ingen inkomst av detta slag (eller uppgift saknas) 1 = Inkomst av detta slag (minst 100 kr)</p>

Variabel	Definition
Socialbidrag	Inkomst av socialbidrag finns i hushållet. 0 = Ingen inkomst av detta slag (eller uppgift saknas) 1 = Inkomst av detta slag (minst 100 kr)
Status	Se sid 38.
Studierande	Studier i folkhögskola, komvux, kvalificerad yrkesutbildning, kompletterande utbildning och/eller högskola. 0 = Inga studier inom dessa skolformer 1 = Studier inom dessa skolformer
Svensk/ utländsk bakgrund	0 = Inrikes födda med minst en inrikes född förälder 1 = Utrikes födda eller inrikes födda med två utrikes födda föräldrar
Sysselsättningsstatus	Sysselsättningsstatus november månad 2005 1 = Sysselsatt 2 = Ej sysselsatt, med kontrolluppgift 3 = Ej sysselsatt, saknar kontrolluppgift
Värnpliktsersättning	Värnpliktsersättning. 0 = Ingen inkomst av detta slag (eller uppgift saknas) 1 = Inkomst av detta slag (minst 100 kr)
Årskurs för avbrott	Årskurs för studieavbrott bland nybörjare i gymnasieskolan höstterminen 2000 1 = Årskurs 1 2 = Årskurs 2 3 = Årskurs 3

Status

En persons status är årsbaserad och klassificeras i sju olika kategorier utifrån individbaserade uppgifter från flera olika datakällor, såsom inkomstuppgifter från kontrolluppgiftssystemet, sysselsättningsregistret vid SCB och Arbetsmarknadsstyrelsens sökanderegister. Kategorierna är följande:

Variabel	Definition
Högskolestuderande	Registrerad på högskolans grundutbildning vår- eller höstterminen och erhållit någon form av studieersättning aktuellt år. Inte klassificerad som etablerad på arbetsmarknaden.
Övriga studerande	Erhållit någon form av studieersättning under aktuellt år. Inte klassificerad som etablerad på arbetsmarknaden eller högskolestuderande.
Värnpliktig	Erhållit värnpliktsersättning under aktuellt år. Inte klassificerad som etablerad på arbetsmarknaden eller studerande.
Etablerad på arbetsmarknaden	Arbetsinkomst på minst 142 100 kronor (år 2005), sysselsatt enligt sysselsättningsregistrets definition (november månad aktuellt år), inga händelser som indikerar arbetslöshet eller arbetsmarknadspolitiska åtgärder som inte är att betrakta som sysselsättning.
Osäker ställning på arbetsmarknaden	<p>- Arbetsinkomst på minst 120 200 kronor och upp till 142 100 kronor (år 2005). Vid inkomst av aktiv näringsverksamhet gäller ingen nedre inkomstgräns. Inte klassificerad som studerande eller värnpliktig.</p> <p>- Arbetsinkomst på minst 142 100 kronor (år 2005) och under året förekomst av arbetslöshet (hel- eller deltid), arbetsmarknadspolitiska åtgärder eller inte klassificerats som sysselsatt enligt sysselsättningsregistrets definition. Inte klassificerad som studerande eller värnpliktig</p>
Svag ställning på arbetsmarknaden	<p>- Arbetsinkomst upp till 120 200 kronor (år 2005). Inte klassificerad som studerande eller värnpliktig.</p> <p>- Arbetsinkomst på minst 120 200 kronor (år 2005) samt under året förekomst av heltidsarbetslöshet/arbetsmarknadspolitiska åtgärder som överstiger 274 dagar. Inte klassificerad som studerande eller värnpliktig.</p>
Utanför arbetsmarknaden	- Avsaknad av arbetsinkomst aktuellt år. Inte klassificerad som värnpliktig eller studerande.

- 2007:1 Några väsentliga sysselsättningsbegrepp i den officiella statistiken
- 2007:2 Registerbaserad aktivitetsstatistik

ISSN 1654-465X (online)
ISSN 1103-7458 (print)

ISBN 978-91-618-1430-5 (print)

Statistikpublikationer kan beställas från SCB, Publikationstjänsten, 701 89 ÖREBRO, e-post: publ@scb.se, telefon: 019-17 68 00, fax: 019-17 64 44. De kan också köpas genom bokhandeln eller direkt hos SCB, Karlavägen 100 i Stockholm. Aktuell publicering redovisas på vår webbplats (www.scb.se). Ytterligare hjälp ges av Information och bibliotek, e-post: information@scb.se, telefon: 08-506 948 01, fax: 08-506 948 99.

Statistical publications can be ordered from Statistics Sweden, Publication Services, SE-701 89 ÖREBRO, Sweden (phone: +46 19 17 68 00, fax: +46 19 17 64 44, e-mail: publ@scb.se). If you do not find the data you need in the publications, please contact Statistics Sweden, Information and Library, Box 24300, SE-104 51 STOCKHOLM, Sweden (e-mail: information@scb.se, phone: +46 8 506 948 01, fax: +46 8 506 948 99).